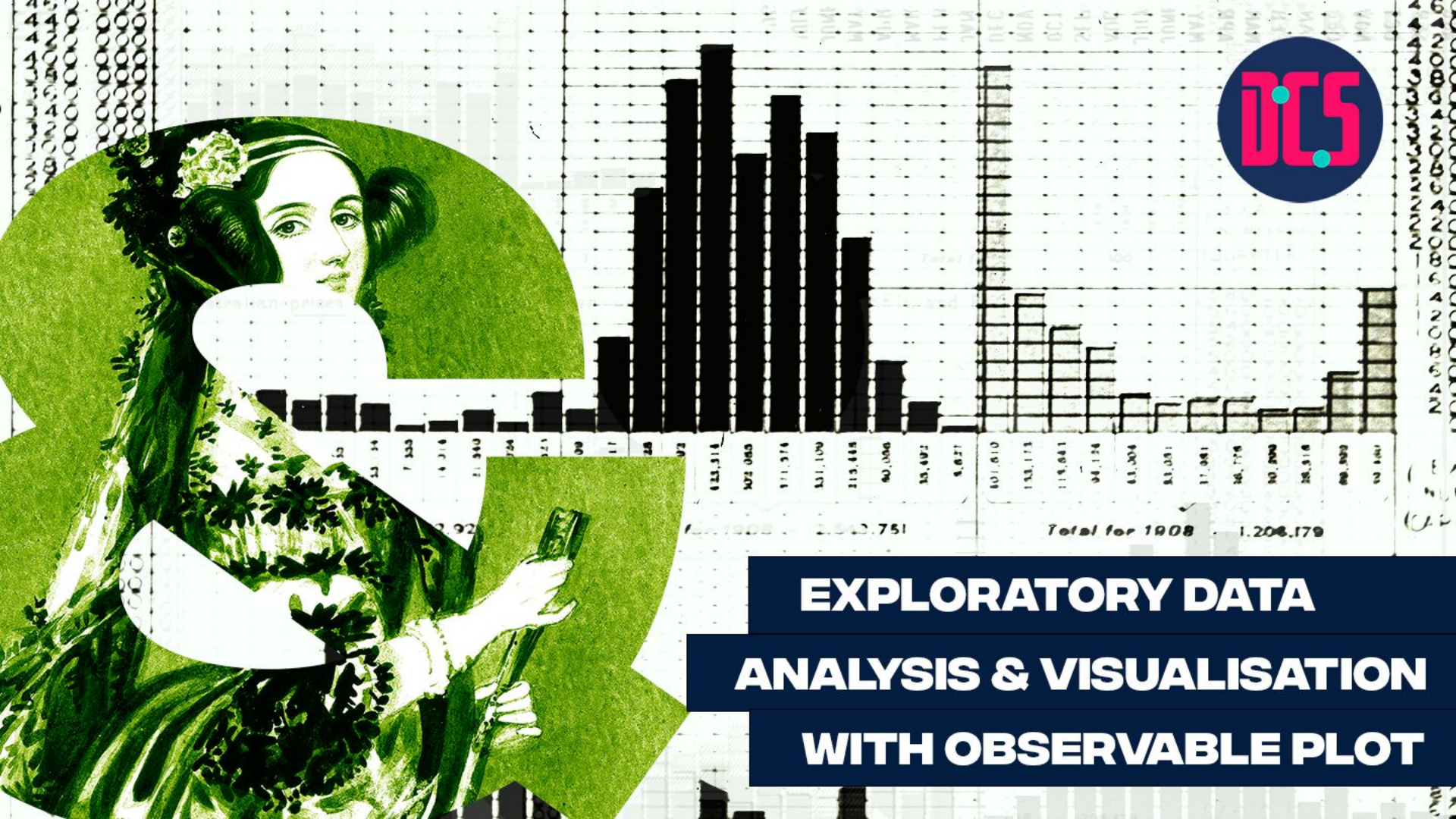

DCS
EXPLORATORY DATA
ANALYSIS & VISUALISATION
WITH OBSERVABLE PLOT

DATA CULTURE SOCIETY

THE UNIVERSITY of EDINBURGH
Edinburgh Futures Institute

VIS HUB University of Edinburgh
Visual+ Interactive Data

Outline

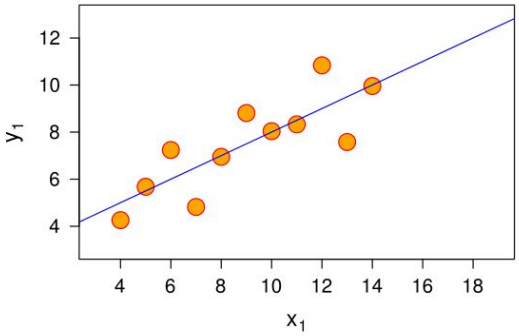
- Introduction
- Grammar of Graphics
- Observable Plot
- Practical Activities

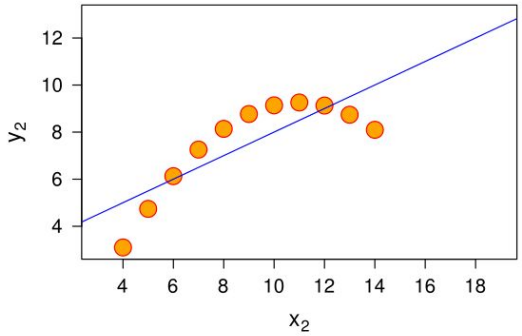
Exploratory Data Analysis

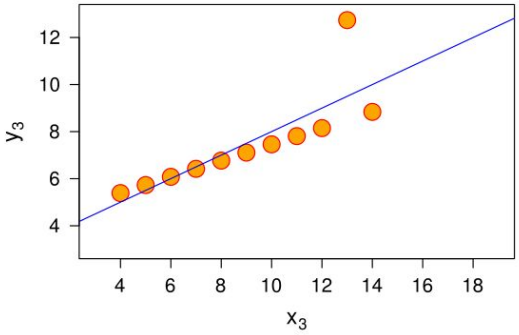
Exploratory data analysis (EDA) is used by data scientists to analyze and investigate data sets and summarize their main characteristics, often **employing data visualization methods**.

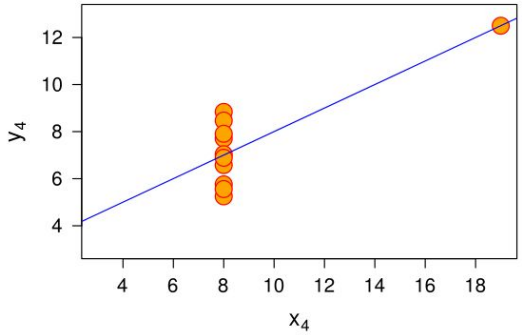
Defined by John W. Tukey in 1977

Anscombe quartet









Property	Value
Mean of x	9
Sample variance of x: s_x^2	11
Mean of y	7.50
Sample variance of y: s_y^2	4.125
Correlation between x and y	0.816
Linear regression line	$y = 3.00 + 0.500x$
Coefficient of determination of the linear regression: R^2	0.67

Many Visualisation Designs…

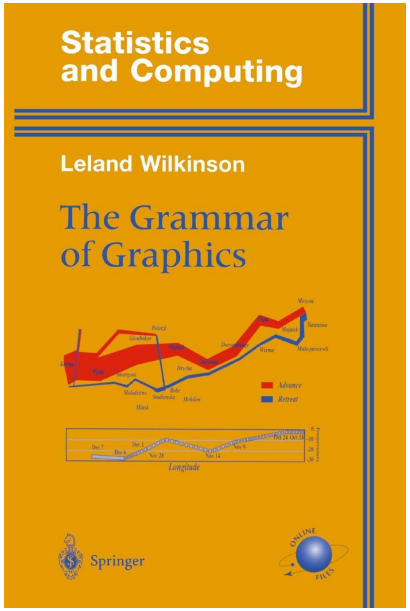
…and many toolkits/ways of doing them

- **Graphical Interface:** Tableau, RawGraph, Flourish
- **Chart-based libraries:** Matplotlib, bokeh, R base library
- **Low-level libraries:** D3, three.js, HTML canvas
- **GoG libraries:** vega, ggplot, Observable Plot

The grammar of graphics

“A grammar of graphics is a tool that enables us to concisely **describe** the **components of a graphic**. Such a grammar allows us to **move beyond named graphics** (e.g., the ``scatterplot'') and gain insight into the deep structure that underlies statistical graphics.”

A Layered Grammar of Graphics, Hadley WICKHAM, 2010



Leland Wilkinson, The Grammar of Graphics, Springer, 1999

The grammar of graphics

- Data
- Transforms (statistics)
- Scales
- Coordinate system
- Geometries (mark)
- Aesthetic mappings
- Facets

The grammar of graphics

Data	
Transforms (statistics)	bin, mean, median, max…
Scales	Linear, log…
Coordinate system	Cartesian, polar…
Geometries (mark)	Points, rect, lines, shapes…
Aesthetic mappings	x, y, fill, stroke…
Facets	

ScatterPlot

ID	culmen length (mm)	culmen depth (mm)
1	39	18
2	38	17
3	28	20

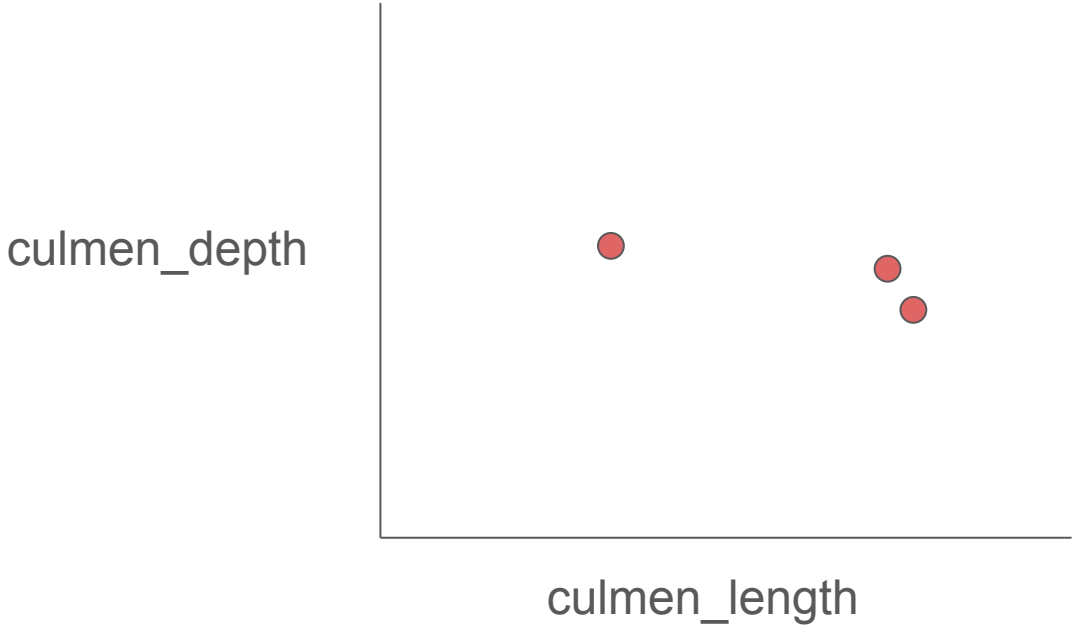
Penguins

ScatterPlot

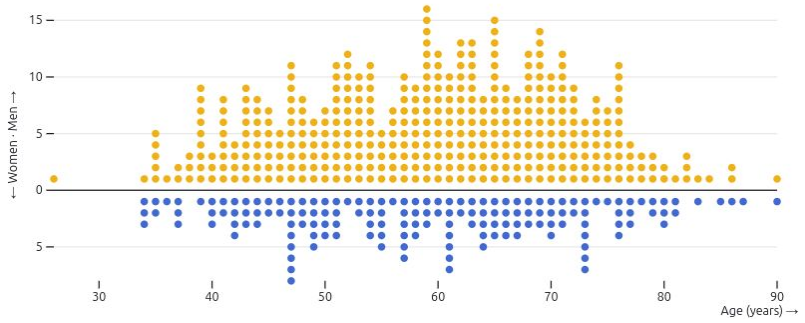
Data	Penguins
Transforms (statistics)	Identity
Scales	Linear
Coordinate system	Cartesian
Geometries (mark)	Points
Aesthetic mappings	x=culmen_length, y=culmen_depth, fill=”red”
Facets	

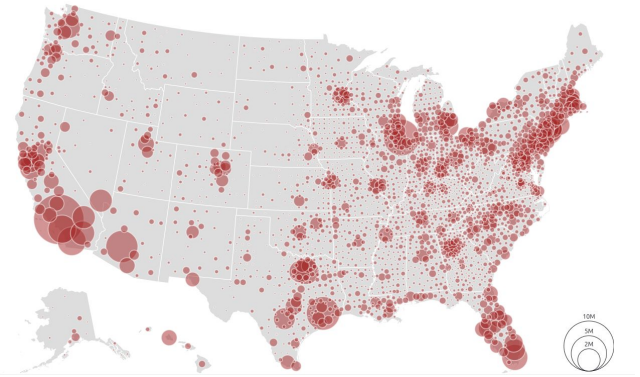
ID	culmen length (mm)	culmen depth (mm)
1	39	18
2	38	17
3	28	20

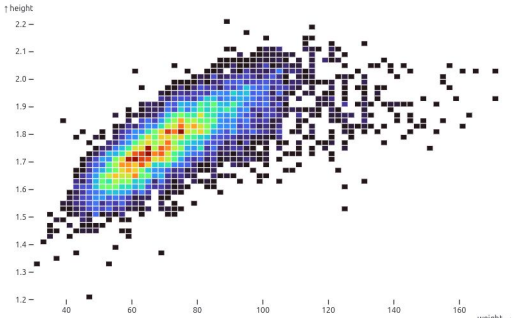
ScatterPlot

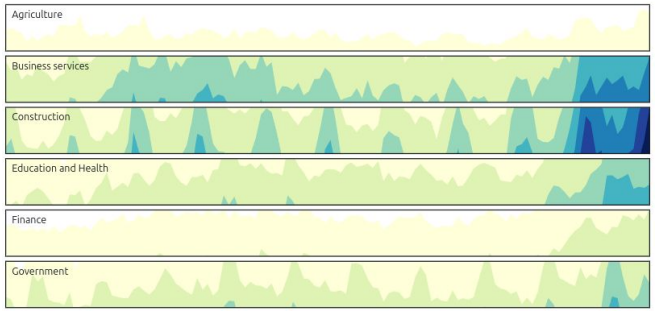


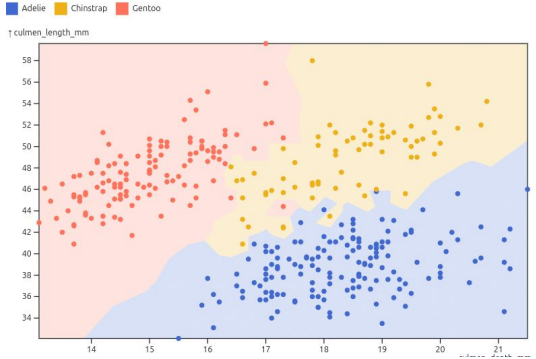
Observable Plot





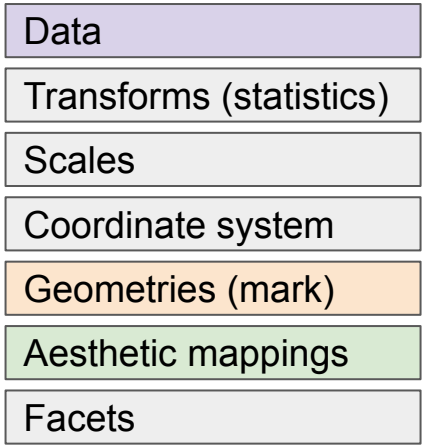






Observable Plot (Scatterplot)

```
Plot.plot({
    marks: [
        Plot.dot(penguins, {x: “culmen_length”, y=”culmen_depth”, fill=”red”})
    ]
})
```



Observable Plot: marks, channels, scales

- Data
- Transforms (statistics)
- Scales
- Coordinate system
- **Geometries (mark)**
- Aesthetic mappings (channel)
- Facets

Plot.dot, Plot.rule, Plot.line, …

Observable Plot: marks, channels, scales

- Data
- Transforms (statistics)
- Scales
- Coordinate system
- Geometries (mark) — Plot.dot, Plot.rule, Plot.line, …
- **Aesthetic mappings (channel)** — {x, y, stroke, fill, strokeWidth, …}
- Facets

Observable Plot: marks, channels, scales

Data	
Transforms (statistics)	
Scales	Plot({x: {type: ‘log’}})
Coordinate system	
Geometries (mark)	Plot.dot, Plot.rule, Plot.line, …
Aesthetic mappings (channel)	{x, y, stroke, fill, strokeWidth, …}
Facets	

Data

ID	culmen length (mm)	culmen depth (mm)
1	39	18
2	38	17
3	28	20

Tabular

```
penguins = [
        {ID: 1, culmen_length: 39, culment_depth: 18},
        {ID: 2, culmen_length: 38, culment_depth: 17},
        {ID: 3, culmen_length: 28, culment_depth: 20}
]
```

JSON

Activities